BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

English translation of DE 36 08 489 for Mrs. Laurie Gathman (Tarrytown).

Abstract

Method of improving image segmentation of a video telephone scene

The transmission bitrate of a video telephone scene is to be reduced by means of image segmentation, in which only the foreground information represented by the moving recording object needs to be transmitted. The essentially stationary background information can be taken from a memory.

For a better separation of foreground information from background information, distance vectors are derived from the known distance-dependent offset of the object contours of two stereoscopically recorded television images in accordance with the methods hitherto known for determining motion vectors. Only distance vectors as from a given value are used for the subsequent image-processing operation. A further accentuation of the recording object can be achieved by means of optical systems sharply imaging only the foreground. The foreground may also be accentuated by masking the out-of-focus of the background (Fig. 3).

Claims

1. A method of improving image segmentation of a video telephone scene into variable foreground information, which is represented by the moving recording object, and essentially stationary background information, which is to be transmitted only once, characterized in that two juxtaposed cameras (10) first stereoscopically pick up two television images of the same object in known manner, in that subsequently, in accordance with a known method of determining motion vectors, each pixel is provided with a distance vector on the basis of the distance-dependent offset of the contours of both images of the recording object, which distance vector is reciprocal to the distance between the camera and the recording object, and in that only the distance vectors as from a given value are used for the subsequent image-processing operation.
2. A method as claimed in claim 1, characterized in that a further accentuation of the recording object is achieved by means of an optical system in the pick-up camera sharply imaging only the foreground (Fig. 3).
3. A method as claimed in claims 1 and 2, characterized in that the foreground is accentuated by masking the out-of-focus of the background.
4. A method as claimed in claim 1, characterized in that the distance vectors are determined in accordance with the block-matching method.
5. A method as claimed in claim 1, characterized in that the distance vectors are determined in accordance with differential methods.

DE 36 08 489.

Description

The invention relates to a method as defined in the precharacterizing part of claim 1.

The Figure elucidates the object of the invention. It shows a typical video telephone scene comprising, in the foreground, the interlocutor's head represented as circles 3 and 4 and moving in the direction indicated by the arrow 5, while – apart from the circumstances dealt with hereinafter – the background 1 remains unchanged. When the head in the foreground moves from the broken-line position 3 to the extended position 4, previously covered areas of the background will become visible. The head's movement thus results in a change of the image in the areas 3 and 4.

Relatively simple segmentation methods allow distinction between stationary backgrounds and area changes. However, there are already segmentation methods in which the moving objects are separated from the background areas that become visible as a result of the movement (Klie J. "Codierung von Fernsehsignalen fur niedrige Ubertragingsbitraten" Thesis, Technical University of Hannover, 1978).

The gray values of the changed image areas are transmitted as relevant information to the receiver: the stationary background – being unchanged from image to image – is taken from a background memory and does not constitute a load for the transmission channel. In order that the receiver inserts the changed and transmitted image areas into the right position in the stationary background again, distance measurements up to an image edge, referred to as addresses, are also transmitted and evaluated by the receiver. This method is known as conditional replenishment.

The data stream loading the channel is thus constituted by the information described by the moving head and the background that becomes visible, and by the address information. To further reduce the resultant data rate, the document cited proposes to further reduce the detail resolution within moving objects by way of low-pass filtering. This is based on the consideration that the human eye easily tolerates the blurred display of moving objects. The method is implemented in such a way that the low-pass filtering operation is speed-controlled, i.e. more details are displayed upon slower movements than upon faster movements.

The method has some drawbacks, leaving the following problems unsolved:

The segmentation into background, into moving and into unchanged image areas is inadequate: difficulties occur when the overall image is changed due to luminance control of the video camera, when motion is simulated due to superimposed background noise, and when deep shadows in the background, detected as image-to-image changes, require transmission of information which is irrelevant to the conversation.

The background structures that become visible constitute a major part of the image information to be transmitted.

If motion estimation methods are used to improve the interpolation of partial images that have been left out, or to improve a prediction algorithm, background structures will cause problems at the contours of moving objects.

It is an object of the invention to obviate these drawbacks by improving the segmentation.

Fig. 2 shows a typical video telephone scene in a plan view. The person 8 conducting the conversation is sitting in front of the pick-up camera 6 and sees his interlocutor on a display unit 7. This typical scene can be unambiguously divided into a foreground (head-shoulder image of the speaking person) 8 and a background 9; it is thus principally possible to also make a distinction with technical means.

The characteristic features defined in the claim realize this distinction. Advantageous implementations of the segmentation method are defined in the dependent claims.

The segmentation method according to the invention has the advantage that an unambiguous identification of pixels associated with the recording object is possible by way of the distance vectors, so that the background parts becoming visible by movement of the recording object can be excluded from the image-processing operation because of their too small distance vectors. The additional special implementation of the optical system of the pick-up camera enhances this effect.

The invention will hereinafter be elucidated by way of 3 Figures. Fig. 1, already described, shows the object of the invention, Fig. 2, also described, shows the typical video telephone scene in a plan view, Fig. 3 shows the camera arrangement for performing the method according to the invention, and Fig. 4 shows the associated block diagram.

Fig. 3 shows a video telephone scene which is similar to that in Fig. 2. It differs in that 2 cameras 10 are juxtaposed, both of which are provided with an optical system which can be automatically focused to the foreground. One of the two cameras only needs to be suitable for black and white but should otherwise have the same optical properties as the other camera. Technical variants of this principal arrangement, e.g. only one pick-up camera with a double objective, are feasible.

When the images of the left and the right camera are caused to register, the contours of the recorded objects show an offset, with those of the foreground being mutually offset to a stronger extent than those of the background. Technically, the offset can be determined by means of the same methods as are also used in motion estimation for computing motion vectors. Essentially, the block-matching and differential methods have emanated therefrom.

In the known displacement estimation method, it is attempted to determine a possible image offset by comparing local image information in the current frame with that of the previous frame. To this end, a trial section (referred to as window) comprising n pixels by m rows is compared with equally large but locally offset windows of the previous frame.

Now it is attempted to find that displacement for which a maximal similarity results. This optimization problem is solved in the block-matching method by using search strategies in a search range (Koga T. et al.: "A 1.5 Mbit/s Interframe Codec with Motion Compensation", Proc. Int Conf. On Commun.D 8,7.1, June 1983 Boston, MA).

In the differential method, however, the local changes of similarity (by an "a-priori" estimation value) are deduced from the location of the optimum (as e.g. in the generally known Newton iteration method). Cafforio C; Rocca F: "The Differential Method for Image Motion Estimation" in Image Sequence Processing and Dynamic Scene Analysis, edited by T.S. Huang, Berlin, Springer Publishing Company, pp. 104-124, 1983.

The length of the vectors is, however, not a measure of speed in this respect, but a distance measure, i.e. the closer an object is to the camera, the longer the vectors found will be. They represent the object contours, and after a suppression of the shorter vectors, the contours will be more prominent.

The automatic focusing ensures that only the depth-of-focus area denoted by the arrow 21 is imaged sharply, whereas the background 9 is recorded out of focus and with little detail resolution.

The block diagram of Fig. 4 shows the principle of the method. Two time-sequential frames of the left channel 11 and the right channel 12 are applied to a change detector 13 and a motion estimator 14. The motion estimation result is applied to an adaptively adjustable threshold 15 whose output conveys a signal for foreground and background distinction. By logic combination of this signal with the output of the change detector 16, the signals are generated in changed image areas 17, in moving objects of the foreground 18 and in the stationary background 19 for the purpose of distinction.

Thus, the use of the motion estimation method is simplified, and the background becoming visible upon motion contributes to the transmitted information to a small extent only because it is recorded out of focus. Also the operations of deriving object contours by convolution with a special operator, and thresholding for suppressing low-value convolution products are simplified.

(Geuen W.: "Konturfindung auf der Basis des visuellen Konturempfindens", Thesis, University of Hannover, 1983.)

The effect of "sharply defined contours of the speaking person and out-of-focus background" and the accentuation of contours can even be enhanced by means of an aperture correction or by out-of-focus masking.

Arp F.: "Normgerechte Aperturkorrektur von Farbfernsehsignalen", NTZ 27 (1974) H.4, pp. 134-138.

In summary, the invention allows separation of the recorded person with the associated objects from the background. This distinction provides the possibility of determining which changed image areas belong to the background. Once transmitted, they are stored in a background memory and when they become visible again in the course of the telephone conversation, they are taken from this memory. Thus, they do not lead to a further load of the transmission channel. In the extreme case, an arbitrary, perhaps locally generated background may be inserted.

3608489

Nummer: 36 08 489
Int. Cl.4: H 04 N 7/12
Anmeldetag: 14. März 1986
Offenlegungstag: 17. September 1987

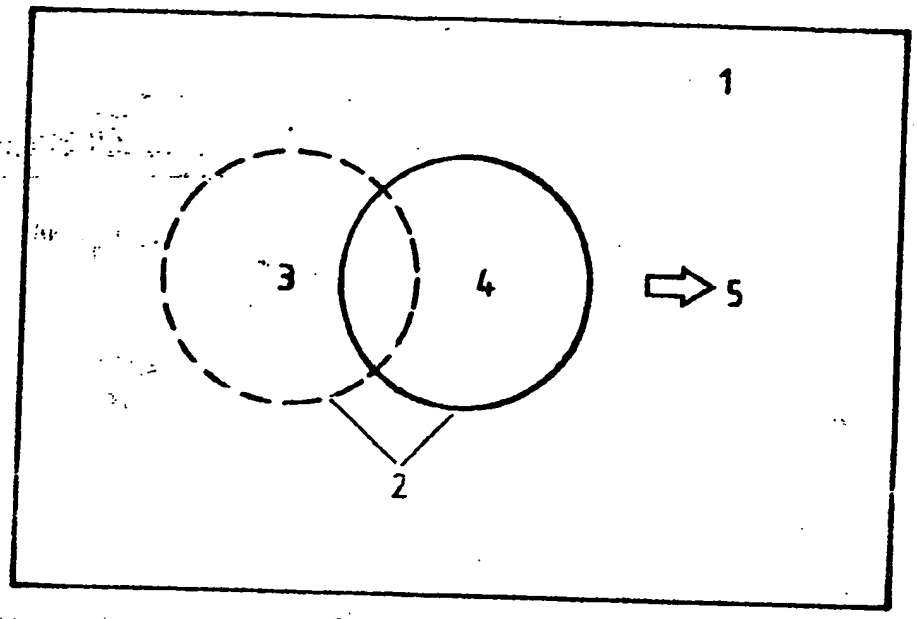


Fig 1

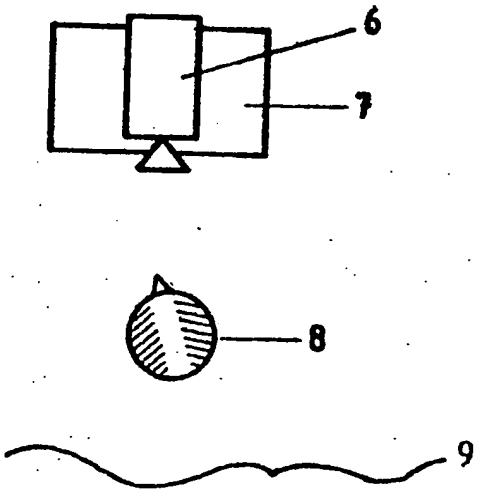


Fig 2

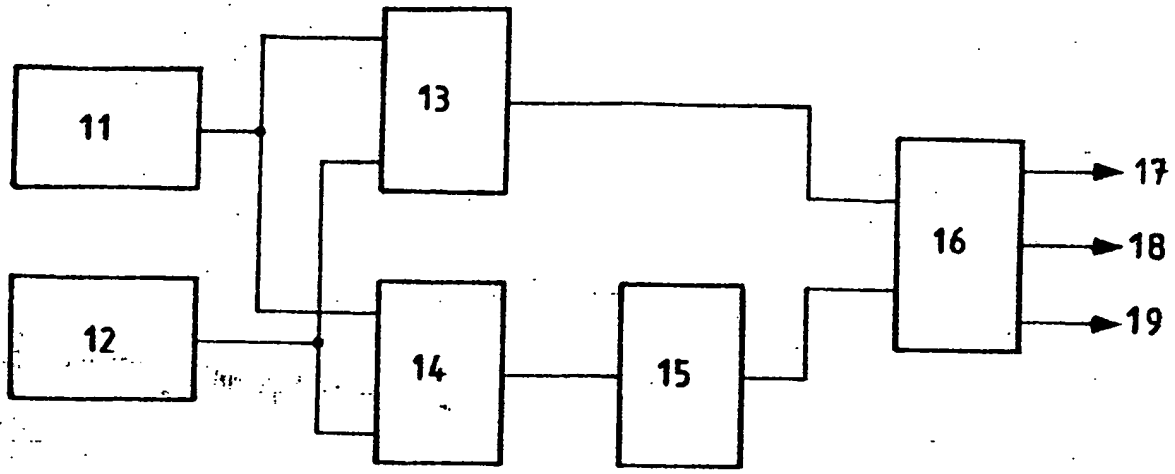


Fig. 4

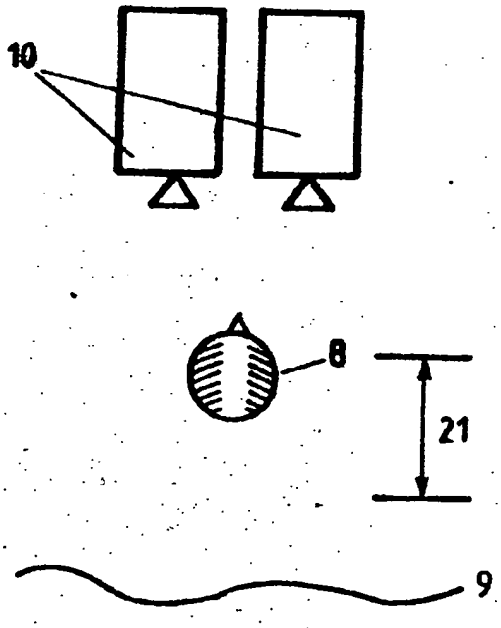


Fig 3

(19) BUNDESREPUBLIK DEUTSCHLAND

DEUTSCHES PATENTAMT

(12) Offenlegungsschrift

(11) DE 3608489 A1

Recherche

(51) Int. Cl. 4:
H04N 7/12
H 04 N 7/14
// H04N 13/00

(21) Aktenzeichen: P 36 08 489.1
(22) Anmeldetag: 14. 3. 86
(43) Offenlegungstag: 17. 9. 87

Reg.

DE 3608489 A1

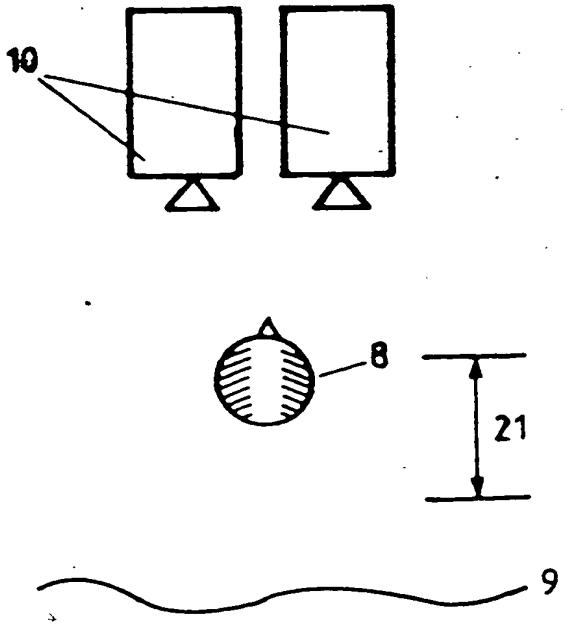
(71) Anmelder:
Robert Bosch GmbH, 7000 Stuttgart, DE

(72) Erfinder:
Stenger, Ludwig, Dr.-Ing., 6451 Mainhausen, DE

(54) Verfahren zur Verbesserung der Bildsegmentierung einer Bildfernsprech-Szene

Die Übertragungsbitrate einer Bildfernsprech-Szene soll durch eine Bildsegmentierung herabgesetzt werden, bei der nur die durch das bewegte Aufnahmeobjekt dargestellte Vordergrundinformation zu übertragen werden braucht. Die im wesentlichen stationäre Hintergrundinformation kann einem Speicher entnommen werden.

Zur besseren Trennung von Vordergrund- und Hintergrundinformation werden aus dem in bekannter Weise entfernungsabhängigen Versatz der Objektkonturen zweier stereoskopisch aufgenommener Fernschilder nach den bisher für die Ermittlung von Bewegungsvektoren verwendeten Verfahren Entfernungsvektoren gewonnen. Für die anschließende Bildverarbeitung werden lediglich Entfernungsvektoren ab einer bestimmten Größe verwendet. Eine weitere Hervorhebung des Aufnahmeobjekts ist durch lediglich den Vordergrund scharf abbildende Optiken erzielbar. Auch kann der Vordergrund durch optische Unschärfemarkierung des Hintergrundes hervorgehoben werden (Fig. 3).



DE 3608489 A1

Patentansprüche

1. Verfahren zur Verbesserung der Bildsegmentierung einer Bildfernsprech-Szene in eine durch das bewegte Aufnahmeobjekt dargestellten variablen Vordergrundinformation und eine im wesentlichen stationäre, nur einmal zu übertragende Hintergrundinformation, **dadurch gekennzeichnet,** daß zunächst in bekannter Weise durch zwei nebeneinander angeordnete Kameras (10) zwei Fernsehbilder des gleichen Objekts stereoskopisch aufgenommen werden, daß darauf aus dem entfernungsabhängigen Versatz der Konturen der beiden Bilder des Aufnahmeobjektes nach einem der für die Ermittlung von Bewegungsvektoren bekannten Verfahren jedes Bildelement mit einem Entfernungsvektor versehen wird, der sich reziprok zur Entfernung der Kamera vom Aufnahmeobjekt verhält, und daß für die anschließende Bildverarbeitung lediglich die Entfernungsvektoren ab einer bestimmten Größe verwendet werden (**Fig. 3**).

2. Verfahren nach Patentanspruch 1, dadurch gekennzeichnet, daß durch eine lediglich den Vordergrund scharf abbildende Optik in den Aufnahmekameras eine zusätzliche Hervorhebung des Aufnahmeobjekts erreicht wird (**Fig. 3**).

3. Verfahren nach den Patentansprüchen 1 und 2, dadurch gekennzeichnet, daß der Vordergrund durch eine Unschärfemaskierung des Hintergrundes hervorgehoben wird.

4. Verfahren nach Patentanspruch 1, dadurch gekennzeichnet, daß die Ermittlung der Entfernungsvektoren nach dem sogenannten Block-Matching-Verfahren erfolgt.

5. Verfahren nach Anspruch 1, dadurch gekennzeichnet, daß die Ermittlung der Entfernungsvektoren mit differentiellen Methoden erfolgt.

Beschreibung

Die Erfindung betrifft ein Verfahren nach dem Gattungsbegriff des Patentanspruchs 1.

Die **Fig. 1** verdeutlicht die Aufgabenstellung. Sie zeigt eine typische Bildfernsprech-Szene, die im Vordergrund den hier als Kreis **3** bzw. **4** dargestellten Kopf des Gesprächspartners enthält, der sich in der durch den Pfeil **5** angedeuteten Richtung bewegt, während der Hintergrund **1** — abgesehen von dem im folgenden behandelten Sachverhalt — unverändert bleibt. Bei der Bewegung des Kopfes im Vordergrund von der gestrichelt gezeichneten Stellung **3** in die ausgezogene Stellung **4** werden vorher abgedeckte Bereiche des Hintergrundes frei. Die Bewegung des Kopfes hat also eine Bildänderung in den Bereichen **3** und **4** zur Folge.

Einfachere Segmentierungsverfahren vermögen zwischen stationärem Hintergrund und geänderten Bereichen zu unterscheiden. Es sind aber auch bereits Segmentierungsverfahren bekannt, welche die bewegten Objekte von den durch die Bewegung frei werdenden Hintergrundbereichen trennen. (Klie, J.: "Codierung von Fernsehsignalen für niedrige Übertragungsbitraten", Dissertation, TU Hannover, 1978).

Die Grauwerte der geänderten Bildbereiche werden als relevante Information zum Empfänger übertragen; der stationäre Hintergrund wird - da von Bild-zu-Bild unverändert — einem Hintergrundspeicher entnommen und verursacht keine Belastung des Übertragungskanals. Damit der Empfänger die geänderten und übertragenen Bildbereiche wieder an der richtigen Stelle in den stationären Hintergrund einfügt, werden Abstandsmaße zu einem Bildrand, sog. Adressen, mitübertragen und vom Empfänger ausgewertet. Dieses Verfahren wird mit Conditional-Replenishment bezeichnet.

Der den Kanal belastende Datenstrom setzt sich also zusammen aus der Information, die das sich bewegende Objekt und den freiwerdenden Hintergrund beschreibt und aus der Adressierungsinformation. Um die resultierende Datenrate weiter zu reduzieren, wird in der zitierten Arbeit vorgeschlagen, die Detailauflösung innerhalb von sich bewegenden Objekten durch Tiefpaßfilterung weiter zu vermindern. Es wird dabei von der Überlegung ausgegangen, daß die unscharfe Wiedergabe bewegter Objekte vom Auge noch am ehesten toleriert wird. Das Verfahren wird so ausgebildet, daß die Tiefpaßfilterung geschwindigkeitsgesteuert wirkt, d. h., daß bei geringerer Bewegung mehr Details wiedergegeben werden als bei heftigerer Bewegung.

Das Verfahren hat einige Schwächen und läßt folgende Probleme ungelöst:

> Die Segmentierung in Hintergrund, in bewegte und in geänderte Bildbereiche ist unvollkommen; so ergeben sich Schwierigkeiten, wenn durch die Helligkeitsregelung von Videokameras im gesamten Bild Änderungen verursacht werden, wenn durch überlagertes Hintergrundrauschen Bewegung vorgetäuscht wird und wenn aufgrund von Schlagschatten im Hintergrund, die als Bild-zu-Bildänderungen erkannt werden, für das Gespräch nicht relevante Information übertragen werden muß.
>
> Die freiwerdenden Hintergrundstrukturen verursachen einen Großteil der zu übertragenden Bildinformation.
>
> Werden, um die Interpolation ausgelassener Teilbilder oder um einen Prädiktionsalgorithmus zu verbessern, Bewegungsschätzverfahren eingesetzt, verursachen Hintergrundstrukturen Probleme an Konturen von sich bewegenden Objekten.

Die Erfindung hat zum Ziel, durch Verbesserung der Segmentierung diese Schwächen zu beseitigen.

In der **Fig. 5** ist eine typische Bildfernsprech-Szene in Draufsicht skizziert. Die das Gespräch führende Person **8** sitzt vor der aufnehmenden Kamera **6** und beobachtet den Gesprächspartner auf einem Sichtgerät **7**. Diese typische Szene ist eindeutig in Vordergrund (Kopf-Schulterbild der sprechenden Person) **8** und Hintergrund **9** einzuteilen; es besteht somit die prinzipielle Möglichkeit, auch mit technischen Mitteln eine Unterscheidung vorzunehmen.

Diese Unterscheidung wird durch die im Patentanspruch angegebenen Merkmale herbeigeführt. Vorteilhafte Ausgestaltungen des Segmentierungsverfahrens sind in den Unteransprüchen angegeben.

Das Segmentierungsverfahren der Erfindung hat den Vorteil, daß eine eindeutige Identifizierung von zum Aufnahmeobjekt gehörenden Bildelementen durch die Entfernungsvektoren möglich ist und damit die durch Bewegung des Aufnahmeobjektes frei werdenden Teile des Hintergrundes aufgrund ihrer zu kleinen Entfernungsvektoren von der Bildverarbeitung ausgeschlossen werden können. Die zusätzlich vorgesehene besondere Ausbildung der Optik der Aufnahmekamera verstärkt diese Wirkung.

Im folgenden wird die Erfindung anhand von 3 Figuren näher erläutert. Es zeigen die bereits besprochene

Fig. 1 die Aufgabenstellung, die Fig. 2 die bereits ebenfalls behandelte typische Bildfernsprech-Szene in Draufsicht, die Fig. 3 die Kameraanordnung zur Durchführung des erfindungsgemäßen Verfahrens, die Fig. 4 das zugehörige Blockschaltbild.

Die Fig. 3 zeigt eine ähnliche Bildfernsprech-Szene wie die Fig. 2; zum Unterschied zu dieser sind jetzt 2 Kameras 10 nebeneinander angeordnet, die beide mit einer automatisch auf den Vordergrund fokussierbaren Optik versehen sind. Eine der beiden Kameras braucht nur schwarz-weißtauglich zu sein, sollte aber im übrigen die gleichen optischen Eigenschaften wie die andere Kamera besitzen. Technische Varianten dieser prinzipiellen Anordnung, z. B. nur eine Aufnahmekamera mit Doppelobjektiv, sind denkbar.

Bringt man die Bilder der linken und der rechten Kamera zur Deckung, so zeigt sich ein Versatz der Konturen der aufgenommenen Objekte, wobei diejenigen des Vordergrundes stärker gegeneinander verschoben sind als die des Hintergrundes. Technisch läßt sich der Versatz mit den gleichen Methoden ermitteln, die auch bei der Bewegungsschätzung zur Berechnung von Bewegungsvektoren eingesetzt werden. Hier haben sich im Wesentlichen die sog. Block-Matching-Verfahren und differentielle Methoden herauskristallisiert.

Beim bekannten sogenannten Displacement-Schätzverfahren (Displacement, englisch für Verschiebung), versucht man einen eventuellen Bildversatz durch Vergleiche lokaler Bildinformationen im aktuellen Bild mit denen des vorangehenden Bildes zu ermitteln. Hierzu wird ein *n* Bildpunkte mal *m* Zeilen umfassender Probeausschnitt (ein sog. Fenster) mit gleichgroßen aber örtlich verschobenen Fenstern des vorhergehenden Bildes verglichen.

Man versucht nun die Verschiebung zu finden, für die sich eine möglichst große Ähnlichkeit ergibt. Dieses Optimierungsproblem wird bei sog. Block-Matching-Verfahren durch Suchstrategien in einem Suchbereich gelöst (Koga, T. et al.: "A 1.5 Mbit/s Interframe-Codec with Motion Compensation", Proc. Int. Conf. on Commun., D 8,7.1, June 1983 Boston, MA).

Bei den differentiellen Verfahren hingegen wird aus den lokalen Veränderungen der Ähnlichkeit (um einen "a-priori"-Schätzwert) auf die Lage des Optimums geschlossen (wie z. B. beim allg. bekannten Newtonschen Iterationsverfahren). Cafforio, C.; Rocca, F.: "The Differential Method for Image Motion Estimation", in Image Sequence Processing and Dynamic Scene Analysis, edited by T. S. Huang, Berlin, Springer-Verlag, pp. 104—124, 1983.

Die Länge der Vektoren ist im vorliegenden Zusammenhang jedoch nicht ein Maß für die Geschwindigkeit, sondern ein Abstandsmaß, d. h. je näher sich ein Objekt zur Kamera befindet, um so länger werden die ermittelten Vektoren. Diese repräsentieren die Objektkonturen und nach einer Unterdrückung der kürzeren Vektoren werden die Konturen deutlicher hervortreten.

Durch die automatische Fokussierung ist sichergestellt, daß lediglich der mit dem Pfeil 21 bezeichnete Schärfentiefebereich scharf abgebildet wird, während der Hintergrund 9 unscharf und mit geringer Detailauflösung aufgenommen wird.

Das Blockschaltbild nach Fig. 4 zeigt das Verfahrensprinzip. Zwei zeitlich aufeinanderfolgende Bilder des linken 11 und des rechten Kanals 12 werden jeweils einem Änderungsdetektor 13 und einem Bewegungsschätzer 14 zugeführt. Das Ergebnis der Bewegungsschätzung wird über eine adaptiv einstellbare Schwelle 15 geleitet, an deren Ausgang ein Signal zur Unterscheidung von Vordergrund- und Hintergrund zur Verfügung steht. Durch logische Verknüpfung dieses Signals mit dem Ausgang des Änderungsdetektors 16 werden die Signale zur Unterscheidung in geänderte Bildbereiche 17, in bewegte Objekte des Vordergrundes 18 und in stationären Hintergrund 19 erzeugt.

Somit wird der Einsatz von Bewegungsschätzverfahren erleichtert, und der bei Bewegung freiwerdende Hintergrund trägt, da unscharf aufgenommen, nur gering zur übertragenen Information bei. Auch die Ableitung von Objektkonturen durch Faltung mit einem speziellen Operator, und Schwellenbildung zur Unterdrükkung geringwertiger Faltungsprodukte wird erleichtert.

(Geuen, W.: "Konturfindung auf der Basis des visuellen Konturempfindens", Dissertation, Univ. Hannover, 1983)

Der Effekt "scharf umrissene Konturen der sprechenden Person und unscharfer Hintergrund" und das Hervorheben von Konturen läßt sich durch eine Aperturkorrektur, bzw. durch eine Unscharfmaskierung noch verstärken.

Arp, F.: "Normgerechte Aperturkorrektur von Farbfernsehsignalen", NTZ 27 (1974) H. 4, S. 134—138.

Insgesamt erlaubt die Erfindung eine Trennung der aufgenommenen Person mit den ihr zugeordneten Einrichtungsgegenständen vom Hintergrund. Durch diese Unterscheidung kann eine Aussage getroffen werden, welche der geänderten Bildbereiche zum Hintergrund gehören. Sie werden, einmal übertragen, in einen Hintergrundspeicher abgelegt und, wenn sie im Verlauf des Gespräches erneut frei werden, diesem entnommen; sie führen somit zu keiner weiteren Belastung des Übertragungskanals. Im Extremfall kann auch ein beliebiger, vielleicht örtlich erzeugter Hintergrund eingeblendet werden.

– Leerseite –

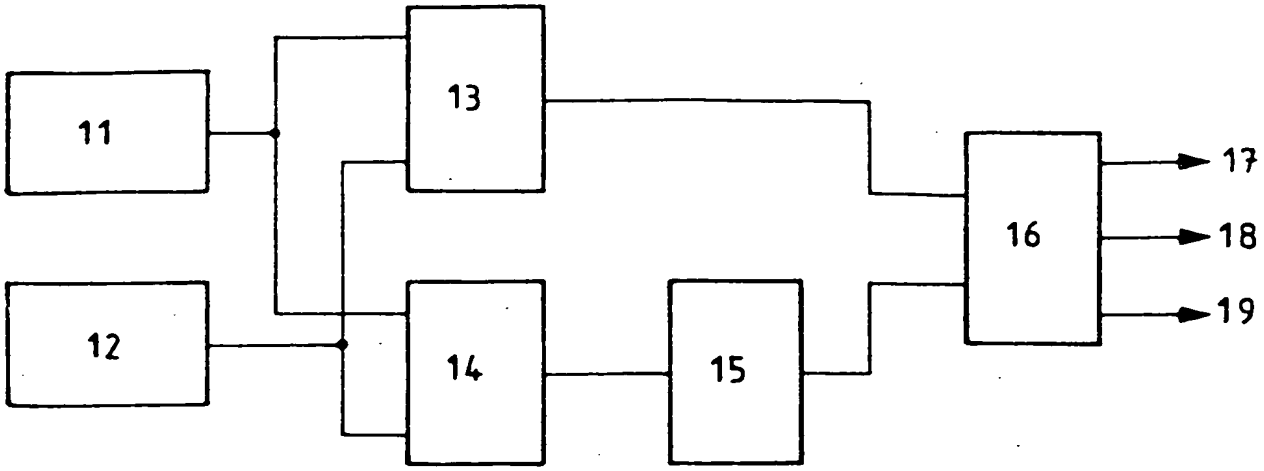


Fig. 4

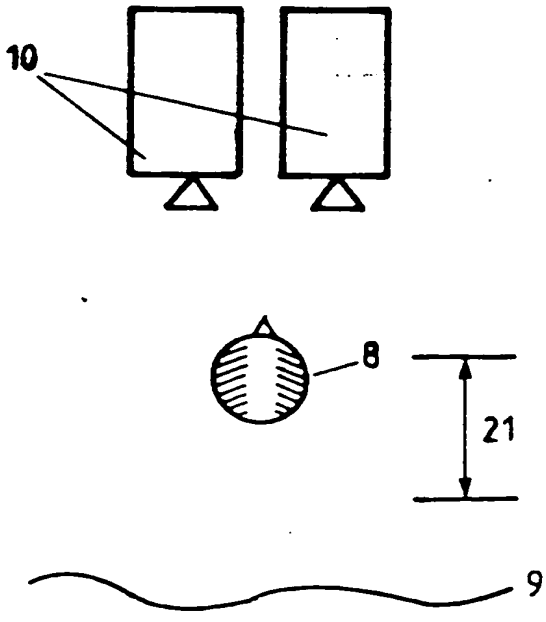


Fig 3

3608489

Nummer: 36 08 489
Int. Cl.⁴: H 04 N 7/12
Anmeldetag: 14. März 1986
Offenlegungstag: 17. September 1987

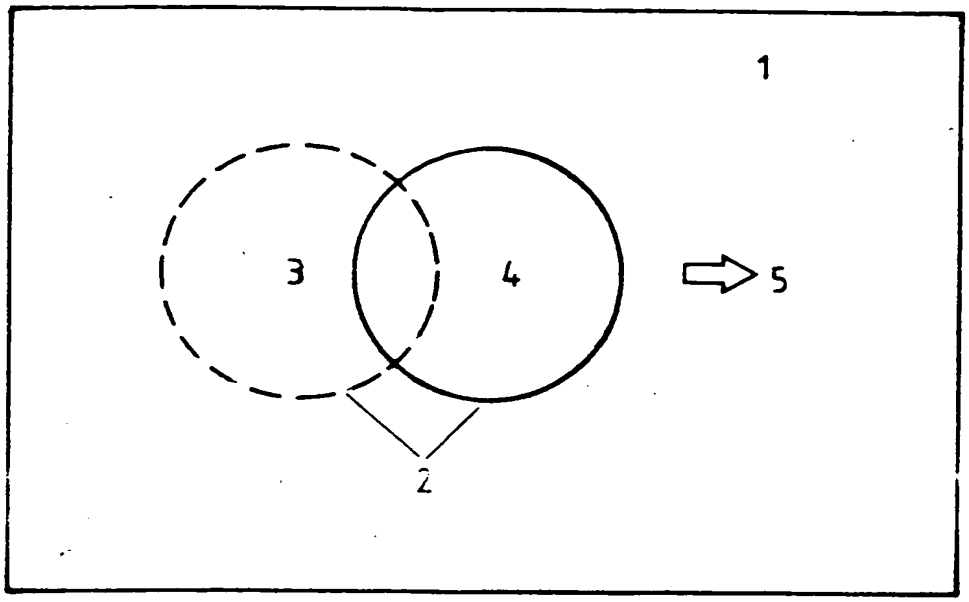


Fig 1

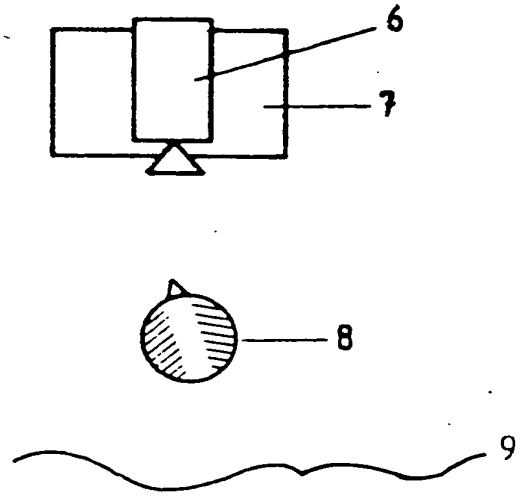


Fig 2